



PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/21731>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.



ELSEVIER

Stochastic Processes and their Applications 57 (1995) 1–10

stochastic
processes
and their
applications

Compactness in the theory of large deviations

George L. O'Brien,^{a,*} Wim Vervaat^b

^a*Department of Mathematics and Statistics, York University, 4700 Keele Street, North York, Ont., Canada
M3J 1P3*

^b*Université Claude Bernard Lyon 1 43 blvd du 11 Novembre 1918, 696 22 Villeurbanne, France*

Received April 1994; revised December 1994

Abstract

Large-deviation principles (LDPs) are expressed as the vague or narrow convergence of sequences of set functions called capacities. Compactness and other topological properties of the collection of capacities are then used in conjunction with Varadhan's integral theorem to reduce the proof of LDPs to the problem of showing that a certain system of equations has a unique solution. As applications of these ideas, we present short proofs of extended versions of a theorem of Bryc and of the Gärtner–Ellis theorem.

This paper was largely written before the recent unexpected death of Wim Vervaat. The probability community has lost a valuable member. I have lost a fine collaborator and a very good friend.—George O'Brien

1. Introduction

Let E be a Hausdorff topological space and assume further that E is second countable or metrizable. Let \mathcal{G} , \mathcal{F} , \mathcal{K} , \mathcal{B} and \mathcal{P} denote the collections of open, closed, compact, Borel and all subsets of E , respectively. For any tight (Borel) probability measure μ and for any $\alpha > 0$, we define μ^α by $\mu^\alpha(A) := (\mu(A))^\alpha$. These assumptions and notations apply throughout this paper.

Definition 1.1. By the term *large-deviation sequence* (LD-sequence), we will mean a sequence of the form $(\mu_n^{\alpha_n})$ where each μ_n is a tight probability measure and where $\alpha_n \rightarrow 0$ in $(0, 1]$.

Definition 1.2. Let $I: E \rightarrow [0, \infty]$ be a lower semi-continuous (l.s.c.) function. The *narrow large-deviation principle* (narrow LDP, NLDP) for the LD-sequence $(\mu_n^{\alpha_n})$ with

* Corresponding author. E-mail: obrien@mathstat.yorku.ca.

rate I is the statement

$$\sup_{x \in G} e^{-I(x)} \leq \liminf \mu_n^{\alpha_n}(G) \quad \text{for all } G \in \mathcal{G}; \quad (1.1)$$

and

$$\sup_{x \in F} e^{-I(x)} \geq \limsup \mu_n^{\alpha_n}(F) \quad \text{for all } F \in \mathcal{F}. \quad (1.2)$$

The vague LDP (VLDP) for (μ_n) with powers (α_n) and rate I is the statement (1.1) and

$$\sup_{x \in K} e^{-I(x)} \geq \limsup \mu_n^{\alpha_n}(K) \quad \text{for all } K \in \mathcal{K}. \quad (1.3)$$

Two general references for large-deviation theory are the books by Deuschel and Stroock (1989) and Dembo and Zeitouni (1993). They refer to NLDPs and VLDPs as “full” and “weak” LDPs, respectively.

A LDP is often viewed as a collection of assertions about sequences of real numbers, one for each G and F (or K). Our approach is to treat a LDP as a single assertion about the convergence of the sequence $(\mu_n^{\alpha_n})$ of set functions to the set function $\sup\{e^{-I(x)}: x \in \cdot\}$, which we denote more simply as e^{-I} . These set functions are all subadditive capacities, in the sense of O'Brien and Vervaat (1991) and O'Brien (1995). Capacities of the form e^{-I} with l.s.c. I are called ‘sup measures’; we denote the set of sup measures by SM . The set of capacities has two topologies of interest, the vague and the narrow topologies. Narrow (vague) LDPs are assertions of convergence of sequences of capacities in the narrow (vague) topology, as defined in the cited papers. This connection allows us to apply general properties of these topologies to the large deviation context. The main consequences for large deviations are listed in Theorem 2.3.

The other major tool of this paper is Varadhan's (1966) integral theorem. Our main results are obtained by a simple amalgamation of these two things. Here is a brief synopsis. Given a LD-sequence, every subsequence has a convergent sub-sequence. Varadhan's theorem then gives us constraints on its limit. If only one $c \in SM$ satisfies all constraints obtained in this way, then $(\mu_n^{\alpha_n})$ must converge to c . The constants we obtain are inequalities or equations: thus the proof of an LDP amounts to the verification that a system of equations or inequalities has a unique solution.

More details of this compactness approach to large deviations are given in Section 2. Two examples of the approach are given in the second half of the paper. In Section 3, we give an extended version of a theorem of Bryc (1990). In Section 5, we prove an extended Gärtner–Ellis theorem. Our compactness arguments permit proofs which are much shorter than the known proofs of these theorems. Of course, some work had to be done to prove the basic compactness results, but now these results can be used in diverse situations. Part of our purpose here is to demonstrate their usefulness.

We remark that Pukhalskii (1991, 1994) has used a similar compactness approach to large deviations.

2. The basic method: Compactness and Varadhan's theorem

Definition 2.1. A LD-sequence is said to be vaguely (narrowly) convergent if there exists a l.s.c. function I such that the vague (narrow) LDP holds with rate I .

Since E is Hausdorff, every narrowly convergent LD-sequence is also vaguely convergent. The reverse implication holds under the following standard additional condition.

Definition 2.2. A LD-sequence $(\mu_n^{\alpha_n})$ is said to be *equitight* if for all $\varepsilon > 0$ there is a compact set K such that $(\mu_n(K^c))^{\alpha_n} < \varepsilon$ for all n . A sup measure c or its rate is said to be *tight* (or *good*) if for all $\varepsilon > 0$ there is a compact set K such that $c(K^c) < \varepsilon$.

Equitightness for capacities extends both the standard notion for probability measures, as described in Billingsley (1968), and the notion of exponential tightness, as used for example in Deuschel and Stroock (1989) and Bryc (1990). In the former case, the narrow topology is usually called the weak topology and equitightness is equivalent to narrow relative compactness, at least on Polish spaces; in general, and in particular in the large-deviation case, this equivalence fails. A general equivalent condition is given in O'Brien (1995), but we will not use it here. We are now ready to list some important convergence properties of LD-sequences.

Theorem 2.3. Let $S := (\mu_n^{\alpha_n})$ be a LD-sequence.

- (a) Then S has a subsequence which is vaguely convergent.
- (b) At most one VLDP can hold for S (that is (1.1) and (1.3) can hold simultaneously for at most one l.s.c. I).
- (c) If S is equitight, then a VLDP for S implies the NLDP with the same rate, which must in this case be tight.
- (d) If S is equitight, then S has a subsequence that is narrowly convergent to a tight sup measure.
- (e) If E is metrizable and S satisfies a NLDP with tight rate c , then S is equitight.

Versions of all these results are well known. Parts (a) and (b) were proved in O'Brien (1995), where also more general capacity-theoretic versions of all the results are given. Part (c) as stated is well known and is given for example in Deuschel and Stroock (1989), as is the NLDP version of (b). Part (d) follows from (a) and (c); slightly weaker versions of (d) were proved using different methods by O'Brien and Vervaat (1991) and by Pukhalskii (1991). Part (e) was proved in Lynch and Sethuraman (1987); its analogue for weak convergence of probability measures is the final theorem of Billingsley (1968).

We now present a version of Varadhan's (1966) integral theorem. Let $(\mu_n^{\alpha_n})$ be a LD-sequence. Let $f: E \rightarrow [-\infty, \infty)$ be continuous. We define the indefinite integral $\mu_{f,n}$ by

$$\mu_{f,n}(A) := \int_A e^{f/\alpha_n} d\mu_n \in [0, \infty], \quad (2.1)$$

where we have suppressed the dependence on α_n in this notation. It turns out that each $\mu_{f,n}^{\alpha_n}$ is also a capacity; we extend the definition of equitightness to these capacities.

Theorem 2.4. *Suppose the LD-sequence $(\mu_n^{\alpha_n})$ satisfies a VLDP with rate I . Let $f: E \rightarrow [-\infty, \infty)$ be continuous.*

(a) *Then*

$$\liminf_{n \rightarrow \infty} \mu_{f,n}^{\alpha_n}(E) \geq \sup_{x \in E} e^{f(x) - I(x)}. \quad (2.2)$$

(b) *If $(\mu_{f,n}^{\alpha_n})$ is equitight, then*

$$\lim_{n \rightarrow \infty} \mu_{f,n}^{\alpha_n}(E) = \sup_{x \in E} e^{f(x) - I(x)}. \quad (2.3)$$

(c) *If $(\mu_n^{\alpha_n})$ is equitight and if, for some $\delta > 1$,*

$$\limsup \left(\int_E e^{\delta f / \alpha_n} d\mu_n \right)^{\alpha_n} < \infty, \quad (2.4)$$

then $(\mu_{f,n}^{\alpha_n})$ is also equitight.

Remarks 2.5. Note that (2.2) and the same assertion with the inequality reversed can be obtained for more general f ; it is sufficient to have appropriately semi-continuous functions which can take the value $+\infty$ with some restrictions. Results of this type are given in Deuschel and Stroock (1989). Other refinements on Varadhan's integral theorem are given in Gerritse (1993). These extensions of Theorem 2.4 can be used to give an alternative proof of Theorem 5.1 below.

At this point, we revert to the more common logarithmic form of LDPs. The next notion extends a concept of Bryc (1990).

Definition 2.6. If $(\mu_n^{\alpha_n})$ is a LD-sequence and $f: E \rightarrow [-\infty, \infty]$ is a measurable function, we say f yields a limit $\psi(f)$ (relative to (μ_n) and (α_n)) if

$$\psi(f) := \lim_{n \rightarrow \infty} \alpha_n \log \int_E e^{f/\alpha_n} d\mu_n \text{ exists in } [-\infty, \infty]. \quad (2.5)$$

The quantity $\psi(f)$ is often called the *pressure* of f .

We now put Theorems 2.3 and 2.4 together to get the central result.

Theorem 2.7. *Let $(\mu_n^{\alpha_n})$ be a LD-sequence and let Λ be a class of continuous $[-\infty, \infty)$ -valued functions on E such that each $f \in \Lambda$ yields a limit $\psi(f)$. Then*

(a) *every vague limit point e^{-I} of $(\mu_n^{\alpha_n})$ satisfies*

$$\bigvee_{x \in E} (f(x) - I(x)) \leq \psi(f) \text{ for all } f \in \Lambda \quad (2.6)$$

and, for those f for which $(\mu_{f,n}^{\alpha_n})$ is equitight, satisfies

$$\bigvee_{x \in E} (f(x) - I(x)) = \psi(f). \quad (2.7)$$

(b) If the system made of (2.6) and (2.7) (of inequalities and equations) has a unique l.s.c. solution I , then $(\mu_n^{\alpha_n})$ converges vaguely to e^{-I} .

(c) If $(\mu_n^{\alpha_n})$ is itself equitight, the convergence is also narrow.

We usually require some specific information about Λ and ψ to apply Theorem 2.7. We can however easily deduce the following general observations from the preceding theorems.

Proposition 2.8. Let (μ_n) , (α_n) and Λ be as in Theorem 2.7.

(a) The system made of (2.6) and (2.7) always has at least one l.s.c. solution I with $e^{-I(x)} \leq \liminf \mu_n^{\alpha_n}(E)$ for all $x \in E$.

(b) If the system (2.6) and (2.7) has a unique such solution I and Λ is enlarged by appending other functions which yield limits then I also satisfies the corresponding enlarged system (2.6) and (2.7).

(c) Every solution I of (2.6) satisfies

$$\bigvee \{f(x) - \psi(f) : f \in \Lambda, f(x) \vee \psi(f) > -\infty\} \leq I(x) \quad \text{for all } x \in E. \quad (2.8)$$

In particular, if $\psi(f) = -\infty$ and $f(x) > -\infty$ for some $f \in \Lambda$ and $x \in E$, then $I(x) = \infty$. If $\psi(f) = -\infty$ for some real valued f , then $I \equiv \infty$ and $(\mu_n^{\alpha_n})$ converges vaguely to the zero capacity.

Note that (2.8) provides an upper bound for every limit point e^{-I} of $(\mu_n^{\alpha_n})$. It extends the upper bound developed by de Acosta (1985). Incidentally, de Acosta also provided a useful sufficient semi-norm condition for equitightness of $(\mu_n^{\alpha_n})$, for topological vector spaces E .

3. A theorem of Bryc

In this section we apply the theorems of Section 2 to prove Theorem T.1.2 and an extension of Theorem T.1.3 of Bryc (1990). Bryc used his theorems as steps towards his innovative proof of an infinite-dimensional version of the Gärtner–Ellis theorem. We note that Bryc's proofs required several pages in all. Our first theorem generalizes his Theorem T.1.3, in that he required $\psi(f) < \infty$ for all $f \in \Lambda$.

Theorem 3.1. Let $(\mu_n^{\alpha_n})$ be an equitight LD-sequence. Let Λ be a class of continuous $[-\infty, \infty)$ -valued functions of E such that (i) each f in Λ yields a limit $\psi(f) \in [-\infty, \infty]$; (ii) every constant function is in Λ ; (iii) if x, y are distinct elements of E and $r > 0$, there is a $f \in \Lambda$ with $f(x) > r$ and $f(y) < 0$; and (iv) Λ is closed under finite pointwise minima. Then $(\mu_n^{\alpha_n})$ satisfies a NLDP with rate I given by

$$I(x) := \sup \{f(x) - \psi(f) : f \in \Lambda, f(x) \vee \psi(f) > -\infty\}, \quad x \in E. \quad (3.1)$$

Also, for those $f \in \Lambda$ which are bounded above,

$$\psi(f) = \sup \{f(x) - I(x) : x \in E\}. \quad (3.2)$$

Proof. Let e^{-I} be any narrow limit point of $(\mu_n^{\alpha_n})$. These exist by Theorem 2.3(d). For each $f \in \mathcal{A}$ and $r > 0$, $f \wedge r \in \mathcal{A}$ by (ii) and (iv). Therefore the hypotheses of the theorem still hold if \mathcal{A} is replaced by $\mathcal{A}' := \{f \in \mathcal{A} : f \text{ bounded above}\}$. By Proposition 2.8(b), it therefore suffices to prove the theorem under the assumption that every $f \in \mathcal{A}$ is bounded above. By Theorem 2.4(c), $(\mu_{f,n}^{\alpha_n})$ is therefore equitight for all $f \in \mathcal{A}$. By Theorem 2.7(a) and Proposition 2.8(c), we then get (3.2) and (2.8). It remains to be shown that for fixed $x \in E$ and $t < I(x)$ there is an $f^* \in \mathcal{A}$ with

$$f^*(x) \vee \psi(f^*) > -\infty \quad \text{and} \quad t \leq f^*(x) - \psi(f^*). \quad (3.3)$$

Since I is l.s.c., there is an open G with $x \in G$ and $I(G) > t$. Also, by Theorem 2.3(d), there is a compact K such that $I(E \setminus K) \geq I(x)$. The constant function $g_0 := t$ already satisfies $g_0 - I \leq 0$ on $G \cup (E \setminus K)$. To deal with $K \setminus G$, use (iii) to obtain for each $y \in K \setminus G$ a $f_y \in \mathcal{A}$ such that $f_y(x) > t$ and $f_y(z) < 0$ for all z in some neighbourhood of y . By the compactness of $K \setminus G$, there are finitely many $f_1, f_2, \dots, f_k \in \mathcal{A}$ such that $f_i(x) > t$ for all i and $g_1 := \min\{f_1, f_2, \dots, f_k\} < 0$ on $K \setminus G$. Since $I \geq 0$, $g_1 - I \leq 0$ on $K \setminus G$. Set $f^* := g_0 \wedge g_1$ to obtain $f^* - I \leq 0$, so that $\psi(f^*) \leq 0$ by (3.2). Since $f^*(x) = t$, we have (3.3). \square

The next result is a special case of Theorem 3.1 although it can also be proved more directly by using Urysohn's lemma to construct a function corresponding to the function f^* in the above proof. It was also proved by Dinwoodie (1993) under the extra assumption that a NLDP holds, and by Bryc (1990) as stated.

Corollary 3.2. *Let E be metrizable, and let $(\mu_n^{\alpha_n})$ be an equitight LD-sequence. Let \mathcal{A} be the class of all bounded continuous \mathbb{R} -valued functions on E . Suppose each f in \mathcal{A} yields a limit $\psi(f) \in [-\infty, \infty]$. Then $\mu_n^{\alpha_n}$ satisfies a NLDP with rate I given by*

$$I(x) = \sup \{f(x) - \psi(f) : f \in \mathcal{A}\}, \quad x \in E. \quad (3.4)$$

Also, (3.2) holds for all $f \in \mathcal{A}$.

4. Convex functions

In this section, we review some needed properties of convex functions. Let $E = \mathbb{R}^d$. We denote the dual of E by E^* as an aid to the reader, although of course $E^* = E$. The following definitions are understood to apply also for the case where E and E^* are switched. A function $f: E \rightarrow [-\infty, \infty]$ is said to be *convex* if $f \equiv -\infty$ or if f is $(-\infty, \infty]$ -valued and

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) \quad \text{for all } x, y \in E, \alpha \in (0, 1). \quad (4.1)$$

Also, f is said to be *strictly convex* if also (4.1) holds with strict inequality whenever $x \neq y$ and the left-hand side of (4.1) is finite. For any $f: E \rightarrow [-\infty, \infty]$, its *convex hull* $\text{conv} f: E \rightarrow [-\infty, \infty]$ is the greatest convex l.s.c. function not exceeding f , and its

convex conjugate or Legendre–Fenchel transform $f^*: E^* \rightarrow [-\infty, \infty]$ is defined by

$$f^*(s) = \bigvee_{x \in E} (\langle s, x \rangle - f(x)). \quad (4.2)$$

Note that f^* is l.s.c. and convex as the supremum of continuous affine functions on E^* . We then have

$$f^* = (\text{conv } f)^*, \quad (4.3)$$

$$f \geq f^{**} := (f^*)^* = \text{conv } f. \quad (4.4)$$

These facts are proved in Rockafellar (1970) and for more general spaces in Ekeland and Temam (1976). Let $D \subset E$ and consider a function $f: D \rightarrow [-\infty, \infty]$. We extend the definitions of $\text{conv } f$ and f^* to such f by interpreting them in terms of the function obtained by extending the domain of f to E and by setting $f(x) = \infty$ for $x \in E \setminus D$. We now apply Theorem 2.7 to the current context.

Theorem 4.1. *Let $(\mu_n^{\alpha_n})$ be a LD-sequence. Assume that s yields a limit $\psi(s) \in [-\infty, \infty]$ for all $s \in E^*$.*

(a) *Then every vague limit point e^{-I} of $(\mu_n^{\alpha_n})$ satisfies*

$$I^* \leq \psi \text{ on } E^* \quad \text{and} \quad I \geq \text{conv } I = I^{**} \geq \psi^*.$$

In particular, if $\psi(s) = -\infty$ for any $s \in E^$ then $I \equiv +\infty$.*

(b) *If $(\mu_{s,n}^{\alpha_n})$ is equitight for every $s \in E^*$ then e^{-I} also satisfies*

$$I^* = \psi \quad \text{and} \quad I^{**} = \psi^*. \quad (4.5)$$

(c) *If I is convex or if I^{**} is strictly convex, then $I = I^{**}$.*

Proof. Parts (a) and (b) follow from Theorem 2.7(a) and (4.4); the first case of part (c) is obvious, since I is l.s.c. With regard to the second, we first prove that $I^{**}(x) = I(x)$ at every interior point x of the set where I^{**} is finite. Suppose instead that $I^{**}(x) < I(x)$ (recall (4.4)). Let g be an affine function such that $g(y) \leq I^{**}(y)$ for all y with equality only at x . Since I is l.s.c., there is a $\delta > 0$ such that $I(y) > g(y) + \delta$ for all y in some neighbourhood N of x . For some $0 < \varepsilon < \delta$, $I^{**}(y) > g(y) + \varepsilon$ for all y not in N , by the strict convexity. Thus, $I > g + \varepsilon$ everywhere, so the same must hold for I^{**} . This is impossible. This argument extends to points in the relative interior, in the sense of Rockafellar (1970). Since I is l.s.c. and I^{**} is convex, we also get equality at other points where the latter function is finite. \square

We now give a sufficient condition for strict convexity from Ellis (1985, p. 224) or Rockafellar (1970, p. 253). Let $g: E^* \rightarrow (-\infty, \infty]$ be a convex function which is finite on a nonempty convex open set $D \subset E^*$. We remark that g is continuous on D and that if g is lower semi-continuous at a boundary point t of D then g is continuous on all line segments from points in D to t . We say g is *essentially smooth* on D if for every such t and every such line segment the derivative of the restriction of g to the line segment (linearly parameterized) exists at interior points and is unbounded near t . (This definition is shown in Rockafellar (1970, pp. 244, 251, 252) to be equivalent to the usual one involving gradients.)

Proposition 4.2. *Let $g: E \rightarrow (-\infty, \infty]$ be convex. If D is an indicated and g^* is essentially smooth on D , then g is strictly convex.*

5. The Gärtner–Ellis theorem

In this section we again take $E = E^* = \mathbb{R}^d$, and let $(\mu_n^{\alpha_n})$ be a LD-sequence. The standard Gärtner–Ellis theorem (cf. Gärtner, 1977; Ellis, 1984) asserts that if

$$\psi(s) := \lim \alpha_n \log \int_E e^{\langle s, \cdot \rangle / \alpha_n} d\mu_n$$

exists in $(-\infty, \infty]$ for all $s \in E^*$, 0 is an interior point of $B := \{s \in E^* : \psi(s) < \infty\}$, and ψ is essentially smooth, then $\mu_n^{\alpha_n} \rightarrow e^{-I}$ narrowly on E where $I = \psi^*$. We extend this result by requiring B to have a non-empty interior but not that $0 \in \text{int } B$, with the consequence that we do not always get (1.2) for *all* closed sets. (Note that $0 \in B$.) It should also be noted that we do not make the usual assumption that ψ is l.s.c. (or even that ψ is defined everywhere on E^*). A variation of the Gärtner–Ellis theorem has been obtained by Baxter and Jain (1993, Theorem 1.21). It can also be proved by our methods.

In addition to generalizing the Gärtner–Ellis theorem, we demonstrate how compactness can be used to simplify the proof. Note that only the first paragraph of the proof of our theorem is needed in the classical case $0 \in \text{int } D$.

Theorem 5.1. *Let $(\mu_n^{\alpha_n})$ be a LD-sequence. Let*

$$\psi_n(s) := \alpha_n \log \int_E e^{\langle s, \cdot \rangle / \alpha_n} d\mu_n. \quad (5.1)$$

Suppose that $\psi(s) := \lim \psi_n(s)$ exists in \mathbb{R} for all s in an non-empty open convex set $D \subset E^$, and that ψ is essentially smooth on D . Then $\mu_n^{\alpha_n} \rightarrow e^{-\psi^*}$ vaguely in E . If $0 \in D$, then $\mu_n^{\alpha_n} \rightarrow e^{-\psi^*}$ narrowly in E .*

Proof. We begin with the case $0 \in D$. By Chebyshev's inequality and the finiteness of ψ around 0, $(\mu_n^{\alpha_n})$ is equitight. By Theorem 2.4(c) and the fact that D is open, $(\mu_{s,n}^{\alpha_n})$ is equitight for all $s \in D$. Let e^{-I} be any narrow limit point of $(\mu_n^{\alpha_n})$. By Theorem 2.7(a),

$$\bigvee_{x \in E} (\langle s, x \rangle - I(x)) = \psi(s), \quad s \in D. \quad (5.2)$$

Thus I^* is a convex l.s.c. extension of $\psi|_D$ to E^* . By the essential smoothness of ψ , $I^*(s) = \infty$ for $s \notin \text{clos } D$. Similarly, $\psi(s) = \infty$ for such s . Thus $I^* = \psi$ except possibly on the boundary of D . Since I^* is l.s.c. and every ψ_n is convex by Hölder's inequality, we have $I^* \leq \psi$ wherever the latter exists. Since I^* is continuous on line segments joining interior and boundary points of D , we see that $I^{**} = \psi^*$. Since ψ^{**} equals ψ and is essentially smooth on D , ψ^* is strictly convex by Proposition 4.2, so that in fact $I = I^{**} = \psi^*$, as required.

We now consider the general case. We need the following background geometry. Let E_∞ denote the compactification of E formed by appending a point at the end of each ray leading from the origin 0 and by topologizing the resulting set so as to make the map $x \mapsto (1 - \|x\|)^{-1}x$ from the closed unit ball to E_∞ a homeomorphism where, when $\|x\| = 1$, $(1 - \|x\|)^{-1}x$ is interpreted to be the point appended to the ray through x . The infinite sphere $S_\infty := E_\infty \setminus E$ is compact. For $s \in E^*$, extend $\langle s, \cdot \rangle$ from E to E_∞ continuously along rays from 0. Finally, for $s \in E^* \setminus \{0\}$, let H_s denote the compact hemisphere $\{x \in S_\infty : \langle s, x \rangle \in \{0, -\infty\}\}$.

For $s \in D$ and $r > 0$, the set $K(s, r) := \{x \in E_\infty : \langle s, x \rangle \leq r\}$ is compact. By the finiteness of ψ on D and Chebyshev's inequality, $\mu_n^{\alpha_n}(E \setminus K(s, r)) \rightarrow 0$ uniformly in n as $r \rightarrow \infty$. Let $\varepsilon > 0$. For every s_i in a countable dense subset of D , choose r_i such that $\mu_n^{\alpha_n}(E \setminus K(s_i, r_i)) < \varepsilon 2^{-i}$ for all n ; then $\mu_n^{\alpha_n}(E \setminus K) < \varepsilon$ for all n where $K := \bigcap_i K(s_i, r_i)$. Note that $S_D := \bigcap_{s \in D} H_s = K \cap S_\infty = \{x \in S_\infty : \langle s, x \rangle = -\infty \text{ for all } s \in D\}$. Extending each μ_n to $E_D := E \cup S_D$ by setting $\mu_n(S_D) = 0$, we conclude that $(\mu_n^{\alpha_n})$ is equitight on E_D and that $\langle s, \cdot \rangle : E_D \rightarrow [-\infty, \infty)$ is continuous for all $s \in D$. As before, $(\mu_{s,n}^{\alpha_n})$ is equitight for all $s \in D$ and

$$\psi(s) = \bigvee_{x \in E_D} (\langle s, x \rangle - I(x)), \quad s \in D,$$

for every narrow limit point e^{-I} on E_D . Since $\langle s, x \rangle = -\infty$ for all $s \in D$ and $x \in S_D$, we again have (5.2) and the fact that (5.2) has the unique l.s.c. solution $I = \psi^*$ (on E , not on E_D). For any vaguely convergent subsequence of $(\mu_n^{\alpha_n})$ on E , with limit e^{-J} , a subsubsequence converges narrowly to e^{-I} on E_D , for some l.s.c. I with $I = \psi^*$ on E . The restrictions to E then converge vaguely to e^{-I} on E , since E is open in E_D . By Theorem 2.3(b), $J = I$ on E , so the full sequence $(\mu_n^{\alpha_n})$ converges vaguely to $e^{-I} = e^{-\psi^*}$ on E . \square

We cannot expect narrow convergence on E in the case $0 \notin D$ but we do get something more than vague convergence; namely we can easily deduce that $\limsup \mu_n^{\alpha_n}(F) \leq e^{-I(F)}$ for all those closed $F \subset E$ which are bounded away from S_D .

We have shown in the above theorem that if $0 \in D$ (so that $S_D = \emptyset$), then $(\mu_n^{\alpha_n})$ is narrowly convergent on $E_D := E \cup S_D$. We now show that this conclusion always holds in one other case, namely the case when S_D is a singleton. The latter condition holds if $d = 1$ and $0 \notin D$ and more generally if D contains an open ball which has 0 as a boundary point. In fact, the narrow convergence follows in this singleton case from the vague convergence already proved, the fact that for any narrow limit point e^{-J} on E_D , the value $1 = \sup\{e^{-J(x)} : x \in E_D\}$ must be attained for some x , and the fact that $e^{-J(x)} < 1$ for all $x \in E$. Here is a more precise formulation of the last fact. The hypothesis (5.3) is weaker than essential smoothness.

Proposition 5.2. *Let E , μ_n , and α_n be as in Theorem 5.1 and assume that ψ , defined near (5.1), satisfies*

$$\lambda = o(\psi(\lambda s)) \quad \text{as } \lambda \rightarrow 0 \text{ in } (0, 1) \quad (5.3)$$

for some $s \in D$. Then, for all $r \in \mathbb{R}$ and all narrow limit points c of (μ_n^a) on E_D ,

$$c(\{x \in E: \langle s, x \rangle > -r\}) < 1. \quad (5.4)$$

Proof. By Proposition 2.8, $c(E) = 0$ if $\psi(t) = -\infty$ for any t , so we may assume $\psi(t) > -\infty$ for all t . Since ψ is convex and finite on $(\lambda s: 0 < \lambda < 1)$ and $\psi(0) = 0$, we see from (5.3) that $\psi(\lambda s) < 0$ for λ near 0. By Chebyshev's inequality, the left-hand side of (5.4) is at most

$$\begin{aligned} c(\{x \in \mathbb{R}^d: \langle s, x \rangle > -r\}) &\leq \limsup \mu_n^a(\{x \in \mathbb{R}^d: \langle s, x \rangle > -r\}) \\ &\leq \exp\{\lambda r + \psi(\lambda s)\} \end{aligned}$$

which is < 1 for small λ , by (5.3). \square

It is possible for (μ_n^a) to satisfy a NLDP even when the conditions of Theorem 5.1 fail. A trivial example is to take μ_n on \mathbb{R} to have $\mu_n(\{0\}) = \mu_n(\{1\}) = \frac{1}{2}$. Then $\mu_n^{1/n} \rightarrow c \in \text{SM}$ narrowly where $c(x) = I_{(0,1)}(x)$. Since $-\log c$ is not convex, this cannot be deduced from the theorem. Actually, ψ is not differentiable at 0. Similar examples can be given where c is not tight: take all μ_n the same and with an atom at every integer.

References

- J.R. Baxter and N.C. Jain, Convexity and compactness in large deviation theory, preprint (1993).
- P. Billingsley, *Convergence of Probability Measures* (Wiley, New York, 1968).
- W. Bryc, Large deviations by the asymptotic value method, in: M.A. Pinsky, ed., *Diffusion Processes and Related Problems in Analysis*, Vol. 1, (Birkhäuser, Boston, 1990) pp. 447–472.
- A. de Acosta, Upper bounds for large deviations of dependent random vectors, *Z. Wahrsch. verw. Geb.* 69 (1985) 551–565.
- A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (A.K. Peters, Wellesley, MA) (formerly published by Jones and Bartlett, Boston, 1993).
- J.-D. Deuschel and D.W. Stroock, *Large Deviations* (Academic Press, Boston, 1989).
- I.H. Dinwoodie, Identifying a large deviation rate function, *Ann. Probab.* 21 (1993) 216–231.
- I. Ekeland and R. Temam, *Convex Analysis and Variational Problems* (North Holland, Amsterdam, 1976).
- R.S. Ellis, Large deviations for a general class of random vectors, *Ann. Probab.* 12 (1984) 1–12.
- R.S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics* (Springer, New York, 1985).
- J. Gärtner, On large deviations from the invariant measure, *Theory Probab. Appl.* 22 (1977) 24–39.
- B. Gerritse, Varadhan's theorem for capacities, Rept. 9347, Department of Mathematics, Catholic University of Nijmegen (Nijmegen, Netherlands, 1993).
- J. Lynch and J. Sethuraman, Large deviations for processes with independent increments, *Ann. Probab.* 15 (1987) 610–627.
- G.L. O'Brien, Sequences of capacities, with connections to large-deviation theory, to appear in: *J. Theoretical Probab.* (1995).
- G.L. O'Brien and W. Vervaat, Capacities, large deviations and loglog laws, in: S. Cambanis, G. Samorodnitsky and M.S. Taqqu, eds., *Stable Processes and Related Topics* (Birkhäuser, Boston, 1991), pp. 43–83.
- A. Pukhalskii, On functional principle of large deviations, in: V. Sazonov and T. Shervashidze, eds., *New Trends in Probability and Statistics* (VSP, Moscow, 1991) pp. 198–218.
- A. Pukhalskii, The method of stochastic exponentials for large deviations, to appear in *Stochastic Process. Appl.* (1994).
- R.T. Rockafellar, *Convex Analysis* (Princeton Univ. Press, Princeton, 1970).
- S.R.S. Varadhan, Asymptotic probabilities and differential equations, *Comm. Pure Appl. Math.* 19 (1966) 261–286.